



Munich Personal RePEc Archive

## **Application of log-linear analysis in marketing research**

Eva Litavcová and Ivana Butoracová Šindleryová

Prešov University, Faculty of Management, Slovakia

15. November 2009

Online at <http://mpra.ub.uni-muenchen.de/20260/>

MPRA Paper No. 20260, posted 27. January 2010 21:22 UTC

# APLIKÁCIA LOG-LINEÁRNEJ ANALÝZY V MARKETINGOVOM VÝSKUME

Litavcová Eva, Butoracová Šindleryová Ivana

## ABSTRACT

The aim of the article is to point out the possibility of application of log-linear analysis within the research of hidden relations of data. We use real data gained from the marketing research, which was based on the identification of educational institution communication policy tools in order to define the perception of further education need in relation to the value of education from the point of motivation.

## ABSTRAKT

Cieľom príspevku bolo ukázať možnosť aplikácie log-lineárnej analýzy pri hľadaní skrytých vzťahov v dátach. Používame reálne dáta z marketingového výskumu, ktorý bol zameraný na identifikáciu nástrojov komunikačnej politiky vzdelávacej organizácie, na zistenie miery vnímania potreby vzdelávania a na definovanie hodnoty vzdelávania z hľadiska motivácie.

## ÚVOD

V spoločenských vedách, akou nepochybne manažment, marketing, ale aj ekonómia je, sú v rámci skúmanej reality často vyvodzované súvislosti prostredníctvom analýzy empirických dát. Vzťahy skryté v dátach umožňujú odhaliť tieto súvislosti a vyvodiť závery potrebné pre prax. Odhalené súvislosti sa potom v kontexte širších teoretických vedomostí považujú za dôkaz tvrdení. Na identifikáciu zákonitostí v hromadných náhodných javoch sú využívané štatistické analýzy. Paleta štatistických metód, ktorú matematická štatistika ponúka, je široká a ich teoretický základ je pomerne zložitý. Najčastejšie využívané metódy sú tie, ktoré bývajú obsahom základných kurzov štatistiky. Log-lineárna analýza, ktorá patrí do množiny metód zaradovaných medzi „advanced“ techniky, je jednou z tých menej užívaných. Dôvodom jej zriedkavého využívania je nielen jej zložitosť, ale aj skutočnosť, že vyžaduje objemnú vzorku dát. Keďže sme takouto vzorkou v rámci riešenia projektu VEGA 1/4638/07 a výskumu realizovaného v rámci Centra excelentnosti CEVKOG disponovali, ukážeme si príklad využitia tejto analýzy v marketingovom výskume. Na modelovom príklade ukážeme, kam môže viesť chybný výber metódy a, alebo nedodržanie jej predpokladov. Ďalej vysvetlíme princíp log-lineárnej analýzy a napokon uvedieme príklady, ktoré sme získali analýzou empirických dát.

## MODELOVÝ PRÍKLAD

Uvažujme **modelový príklad 1**: Majme dátový súbor troch binárnych premenných  $X$ ,  $Y$ , a  $Z$ , nadobúdajúcich hodnoty 1 a 2. Napríklad, predstavme si, že na burze sledujeme správanie sa troch menových párov. U každého z nich v každom čase nás zaujíma iba stav z hľadiska poklesu (hodnota 1), alebo nárastu (hodnota 2) oproti predchádzajúcemu času. Zaujíma nás, či správanie sa skúmaných troch menových párov vzájomne súvisí. S premennou čas v tomto modeli ďalej nepracujeme. Celkový počet trojíc pozorovaní nech je 24 (Tab. 1). Ich početnosti pri hľadaní súvislosti dvojíc premenných vedú vo všetkých troch prípadoch ku kontingenčným tabuľkám (Tab. 2.), v ktorých sa očakávané početnosti skutočným

početnostiam rovnajú. Jedná sa o ukážkový príklad nesúvisiacich dvojíc premenných, kde zakaždým chí-kvadrát test, aj test pomeru pravdepodobností vedie ku hladine významnosti rovnej 1.

Tab. 1. Dátový súbor

X	1	2	2	2	2	2	1	1	1	2	2	2	1	1	1	2	2	2	1	2	2	2	2	2	n(1)=8, n(2)=16
Y	1	1	1	1	1	1	2	2	2	2	2	2	1	1	1	1	1	1	2	2	2	2	2	2	n(1)=12, n(2)=12
Z	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	n(1)=12, n(2)=12

Zdroj: vlastný príklad

Tab. 2. Skutočné = očakávané početnosti

	Y	1	2		Z	1	2		Z	1	2
X	1	4	4	X	1	4	4	Y	1	6	6
	2	8	8		2	8	8		2	6	6

Tab. 3. Skutočné a očakávané početnosti

	X=1				X=2		
	Z	1	2		Z	1	2
Y	1	1	3	Y	1	2	2
	2	3	1		2	2	2
					Z	1	2
				Y	1	5	3
					2	3	5
				Y	1	4	4
					2	4	4

Zobrazením kontingenčných tabuliek YZ na rôznych úrovniach premennej X (Tab. 3.) vybadáme určitý posun oproti očakávaným početnostiam, ktorý však nevedie k významnosti ( $G^2=2,093$ ;  $p=0,148$  pre  $X=1$ ;  $G^2=1,011$ ;  $p=0,315$  pre  $X=2$ ). Pre  $X=1$  je to posun smerom k negatívnej, pre  $X=2$  smerom k pozitívnej asociácii. Tento jemný posun analýza dvojíc premenných neodhalila a ponechala nás v mylnom dojme, že máme perfektnú nezávislosť troch premenných.

Pri hľadaní súvislostí v tejto vzorke dát ako kompaktného celku trojice premenných, použitím hierarchickej log-lineárnej analýzy dostávame porovnateľný výsledok (Tab. 4.). Model, ktorý najúspornejšie a najlepšie vysvetľuje variabilitu vo vzorke dát je model žiadnej interakcie troch premenných, žiadnej súvislosti dvojíc premenných, ba dokonca sa v modeli nevyskytuje ani hlavný efekt jednotlivých premenných. Ostáva iba generálny efekt, takzvaný Grand Mean model. Najlepším modelom, ktorý vysvetľuje rozloženie početností v tejto vzorke dát je model odhadu očakávaných početností podielom celkového počtu meraní (24) počtom buniek (8).

Tab. 4. Algorithmus Step Summary, tabuľka je neúplná

Step <sup>a</sup>		Effects	Chi-Square <sup>c</sup>	df	Sig.	Number of Iterations
0	Generating Class <sup>b</sup>	X*Y*Z	,000	0	.	
	Deleted Effect	1	3,104	1	,078	2
1	Generating Class <sup>b</sup>	X*Y, X*Z, Y*Z	3,104	1	,078	
	Deleted Effect	1	,000	1	1,000	2
atd...						
6	Generating Class <sup>b</sup>	X	3,104	6	,796	
	Deleted Effect	1	2,718	1	,099	0
7	Generating Class <sup>b</sup>	Constant only	5,822	7	,561	
8	Generating Class <sup>b</sup>	Constant only	5,822	7	,561	

Zdroj: vlastné spracovanie v SPSS

Analýza reziduálnych početností potvrdzuje vhodnosť odhadnutého modelu. Všetky očakávané početnosti optimálneho modelu sú rovné 3 a maximálne štandardizované rezíduá boli, v absolútnej hodnote, rovné 1,155. Z hodnôt štandardizovaných rezíduí je zrejmé, že určitý, i keď nie významný, posun očakávaných početností oproti skutočnosti v nájdenom optimálnom modeli existuje. Napriek rovnakému výsledku z hľadiska významnosti, táto informácia je, pri komplexnom prístupe k trojici premenných, získaná navyše oproti pôvodnej analýze dvojíc premenných.

Uvažujme **modelový príklad 2**. Pre nesplnenie predpokladov použitia niektorých z chí-kvadrát testov v modelovom príklade 1 a aj predpokladov log-lineárnej analýzy, pôvodný modelový dátový súbor skopírujeme, čím získame dvojnásobný počet hodnôt. Výsledok log-lineárnej analýzy, opäť pri neotrasiteľnej nezávislosti dvojíc premenných, sa rapídne zmení. Najoptimálnejším modelom vysvetľujúcim variabilitu vo vzorke dát je teraz model saturovaný, teda najširší model interakcie troch premenných pri 0 stupňoch voľnosti. V modeli je významný efekt interakcie troch premenných  $X*Y*Z$  ( $G^2=6,207$ ;  $p=0,013$ , d.f.=1, tabuľky tu neuvádzame). Najoptimálnejším odhadom očakávaných početností sú ich skutočné početnosti. Každý iný pokus o odhad sa od skutočnosti významne odchyľuje.

Hľadáme dôvody, prečo je to tak. Z tabuľky 5. je zrejmé, že súvislosť  $Y$  a  $Z$  sa môže významne meniť v závislosti od úrovne premennej  $X$ . Pre úroveň premennej  $X=1$  je súvislosť  $Y$  a  $Z$  významná ( $G^2=4,186$ ;  $p=0,041$ ), kdežto pre úroveň  $X=2$  to tak nie je ( $G^2=2,021$ ;  $p=0,155$ ), avšak hladina významnosti aj tu ostáva oveľa menšia ako 1. Súvislosť existovala aj v modelovom príklade 1 (keďže pôvodné dáta sú iba zdvojnásobené), ale predtým malý rozsah skúmaného súboru, ktorý viedol k nesplneniu predpokladov log-lineárnej analýzy, túto skutočnosť nedovolil potvrdiť.

Tab. 5. Skutočné a očakávané početnosti

X=1						X=2					
Z			Z			Z			Z		
1	2		1	2		1	2		1	2	
Y	1	2	6	Y	1	4	4	Y	1	8	8
	2	6	2		2	4	4		2	8	8

Aj pri skúmaní súvislostí  $X$  a  $Z$  a tiež  $X$  a  $Y$  pri rôznych úrovniach tretej premennej sme zistili, že hladina významnosti v každej zo zvyšných tu neopísaných situácií je rovná 0,078 čo je veľmi veľký rozdiel oproti pôvodnej hladine významnosti rovné 1 u každej z dvojíc premenných pri opomenutí vplyvu tretej premennej.

Modelový príklad odhalil, akých omylov sa pri skúmaní dvojíc premenných môže výskumník dopustiť neberúc v úvahu vplyv ďalších premenných a akých omylov sa dopustí, ak neberie do úvahy nedodržanie predpokladov použitej metódy.

## HIERARCHICKÁ LOG-LINEÁRNA ANALÝZA

Ak máme dve kvalitatívne premenné, môže nastať jedna z dvoch situácií:  $X$  a  $Y$  sú nezávislé, alebo sú závislé. Za predpokladu nezávislosti pravdepodobnosť nastania dvoch javov súčasne sa rovná súčinu ich pravdepodobností. Z toho vyplýva, že očakávané bunkové početnosti vypočítavame ako súčin marginálnych početností  $i$ -tej úrovne premennej  $X$  cez všetky úrovne premennej  $Y$  a  $j$ -tej úrovne premennej  $Y$  cez všetky úrovne premennej  $X$ , podelený celkovým počtom meraní.

S navyšovaním počtu premenných dramaticky rastie počet možných situácií. Každú z týchto teoreticky možných situácií nazývame model. Pre každý model je za predpokladu, že daný model je platný, odvodená pravdepodobnostná funkcia vedúca

k výpočtu očakávaných početností, počet stupňov voľnosti a pre test vhodnosti modelu je odvodená testovacia štatistika  $G^2$ , ktorá má chí-kvadrát rozdelenie.

Pri hierarchickej log-lineárnej analýze ide o hľadanie takého modelu súvislosti troch a viacerých kvalitatívnych premenných, ktorý sa hodí na dané dáta a zároveň čo najúspornejšie vysvetľuje vzťahy vo vzorke dát. Algoritmus implicitne postupuje od najzložitejšieho k najoptimálnejšiemu modelu. Model sa hodí na dané dáta, ak štatistika  $G^2$  nie je významná, teda ak odhadnuté očakávané početnosti vyplývajúce z modelu sa významne nelíšia od početností empirických. Ak máme dva modely hodiace sa na dané dáta, vyberieme ten z nich, ktorý je úspornejší, vhodnejší na interpretáciu. Za úspornejší model sa považuje ten, ktorý má menší počet členov (efektov). Pri hierarchickej log-lineárnej analýze začíname najkomplexnejším modelom, tzv. saturovaným modelom, ktorý v sebe zahŕňa všetky možné vzťahy, počet stupňov voľnosti je 0 a očakávané bunkové početnosti sa rovnajú skutočným. Zo saturovaného modelu postupne odoberáme členy, najzložitejšími počnúc, a porovnávame, či po odobratí člena nedošlo k významnému navýšeniu štatistiky  $G^2$ . Kritériom pre rozhodnutie je významnosť rozdielovej štatistiky, teda významnosť príspevku z modelu odoberaného člena.

Pri skúmaní troch premenných pri hierarchickej log-lineárnej analýze vyšetrujeme minimálne tieto modely:

Tab. 6.

Model	Označenie
$\ln(m_{ijk}) = u + u_{X(i)} + u_{Y(j)} + u_{Z(k)}$	(X, Y, Z)
$\ln(m_{ijk}) = u + u_{X(i)} + u_{Y(j)} + u_{Z(k)} + u_{XY(ij)}$	(XY, Z)
$\ln(m_{ijk}) = u + u_{X(i)} + u_{Y(j)} + u_{Z(k)} + u_{XZ(ik)}$	(XZ, Y)
$\ln(m_{ijk}) = u + u_{X(i)} + u_{Y(j)} + u_{Z(k)} + u_{YZ(jk)}$	(YZ, X)
$\ln(m_{ijk}) = u + u_{X(i)} + u_{Y(j)} + u_{Z(k)} + u_{XY(ij)} + u_{YZ(jk)}$	(XY, YZ)
$\ln(m_{ijk}) = u + u_{X(i)} + u_{Y(j)} + u_{Z(k)} + u_{XY(ij)} + u_{XZ(ik)}$	(XY, XZ)
$\ln(m_{ijk}) = u + u_{X(i)} + u_{Y(j)} + u_{Z(k)} + u_{XZ(ik)} + u_{YZ(jk)}$	(XZ, YZ)
$\ln(m_{ijk}) = u + u_{X(i)} + u_{Y(j)} + u_{Z(k)} + u_{XY(ij)} + u_{XZ(ik)} + u_{YZ(jk)}$	(XY, XZ, YZ)
$\ln(m_{ijk}) = u + u_{X(i)} + u_{Y(j)} + u_{Z(k)} + u_{XY(ij)} + u_{XZ(ik)} + u_{YZ(jk)} + u_{XYZ(ijk)}$	(X Y Z)

Zdroj: viac Cristante, 2000

Označenie XY znamená podmiennečnú závislosť premenných X a Y v modeli. Napríklad, model (XY,Z) znamená, že X je podmiennečne nezávislé so Z pri danom Y (pre každú úroveň premennej Y), Y je podmiennečne nezávislé so Z pri danom X, a X a Y sú podmiennečne závislé, pričom asociácie medzi X a Y sú rovnaké pre každú úroveň premennej Z.

Hypotézou pre pravdepodobnosť  $P(X=i, Y=j, Z=k)$  v modeli (XY,Z) je:

$$P_{ijk} = P_{ij} \cdot P_{..k}$$

Minimálnou postačujúcou štatistikou pre odhad očakávaných bunkových početností v modeli sú marginálne rozdelenia početností zodpovedajúce vzťahom v modeli. Pre model (XY,Z) sú to marginálne početnosti  $\{n_{ij.}\}, \{n_{..k}\}$ , pričom  $n_{ij.}$  je marginálna početnosť  $i$ -tej úrovne premennej X a  $j$ -tej úrovne premennej Y cez všetky úrovne premennej Z a  $n_{..k}$  je marginálna početnosť  $k$ -tej úrovne premennej Z cez všetky úrovne premennej X a všetky úrovne premennej Y (Agresti, 2002, str. 334). Očakávaná početnosť pre  $i$ -tú úroveň premennej X,  $j$ -tú úroveň premennej Y a  $k$ -tú úroveň premennej Z v tomto modeli je:

$$m_{ijk} = \frac{n_{ij.} n_{..k}}{N}.$$

Log-lineárna analýza využíva vlastnosť logaritmovania súčinu. Logaritmus očakávanej bunkovej početnosti je vďaka tejto vlastnosti lineárnou funkciou logaritmov jej komponentov. Napríklad pre model  $(XY,Z)$  výraz:

$$\ln(m_{ijk}) = u + u_{X(i)} + u_{Y(j)} + u_{Z(k)} + u_{XY(ij)}$$

z tabuľky 6 znamená, že logaritmus očakávanej bunkovej početnosti je vyjadrený ako súčet príspevku (efektu) celkového priemeru, efektu každej z premenných samostatne a efektu dvojfaktorovej interakcie premenných  $X$  a  $Y$ . Je to súčet celkového priemeru logaritmov početností  $u$ , odchýlky logaritmu očakávanej početnosti  $i$ -tej hodnoty premennej  $X$  od  $u$ , takisto pre  $j$ -tú hodnotu premennej  $Y$ ,  $k$ -tú hodnotu premennej  $Z$  a odchýlky interakcie  $i$ -tej a  $j$ -tej hodnoty premenných  $X$  a  $Y$  od  $u$ . Test pomeru pravdepodobností  $G^2$  má vo všeobecnosti tvar:

$$G^2 = 2 \sum \text{empirická} \cdot \ln \left( \frac{\text{empirická}}{\text{očakávaná}} \right)$$

Pre model  $(XY,Z)$  odvodená testovacia štatistika  $G^2$  má tvar:

$$G_{XY}^2 = 2 \sum_{i=1}^r \sum_{j=1}^c n_{ij.} \ln \left( \frac{N n_{ij.}}{n_{i..} n_{.j.}} \right).$$

Ak sa pri hierarchickej loglineárnej analýze zistí, že sa napr. model  $(XY,Z)$  hodí na dané dáta, (štatistika  $G^2$  nie je významná, model dobre popisuje vzorku v dátach) algoritmus pokračuje odobratím najzložitejšieho efektu z modelu, tu je to efekt  $XY$ , a overením, či týmto nedôjde, pri o jeden stupeň nižšom modeli  $(X,Y,Z)$ , k významnému navýšeniu novej štatistiky  $G^2$ , platnej pre model  $(X,Y,Z)$ . Rozdiel pôvodnej a novej štatistiky  $G^2$  je príspevok efektu  $XY$  a pre prijatie jednoduchšieho modelu nesmie byť významný (algoritmus Step Summary). Algoritmus vyhodnocuje významnosť tejto rozdielovej štatistiky, ktorá má chí-kvadrát rozdelenie s počtom stupňov voľnosti rovnajúcim sa rozdielu d.f. novej – d.f. pôvodnej štatistiky. (Viac Cristante, 2000; Agresti 2002).

Existuje určitá paralela medzi mnohonásobnou regresiou a log-lineárnou analýzou, aj určité rozdiely. Spoločným prvkom je hľadanie komplexných súvislostí vo viac ako dvojrozmernej vzorke dát. Pri regresii hľadáme funkciu najlepšie popisujúcu vzťah medzi vysvetľovanou premennou a vysvetľujúcimi premennými, pri log-lineárnej analýze hľadáme najlepšiu pravdepodobnostnú funkciu pre odhad bunkových početností. Na rozdiel od regresie log-lineárna analýza nepotrebuje vymedziť závisle premennú, nemá veľké nároky na predpoklady a je vhodným nástrojom pre analýzu nominálnych dát. Predpokladom, ktorý tu musíme mať na zreteli je, že každá očakávaná bunková početnosť musí byť  $>1$  a nie viac ako 20% očakávaných početností môže byť  $<5$ . Nedodržanie predpokladu by viedlo k zníženiu sily testu (ukážkou je modelový príklad 1). Keďže počet buniek je rovný súčinu počtov možných úrovní každej z uvažovaných premenných, predpoklad sa dá zabezpečiť iba pri väčších vzorkách dát, kde každá z kombinácií možných hodnôt uvažovaných premenných je reálne možná a naplnená dátami. Ak sa v dátach vyskytujú „deravé bunky“, je možné docielenie splnenia predpokladu spojením susediacich a významovo dobre súvisiacich hodnôt niektorej z premenných do jednej hodnoty. Proces opakujeme dovtedy, kým to možné je. Ako uvidíme v ďalšej časti, na hodnotovej stupnici Likertovej škály pri zisťovaní názorov na určité otázky môže

mať dáta tak silne vychýlené smerom ku kladným alebo záporným postojom, že je vylúčené analyzovať takéto premenné log-lineárnou analýzou.

## **LOG-LINEÁRNA ANALÝZA EMPIRICKÝCH DÁT V MARKETINGOVOM VÝSKUME**

V rámci riešenia projektu sme dotazníkovým prieskumom zozbierali dáta. Dotazník obsahuje celkovo 50 otázok, sedem identifikačných otázok a štyridsaťtri škálových, s identickou škálou v rozpätí 1-5 (rozhodne nie – rozhodne áno), ktoré sú rozdelené do štyroch hlavných oblastí. Identifikačné otázky tvoria prvú časť dotazníka. Druhá časť obsahuje otázky zamerané na identifikáciu nástrojov komunikačnej politiky vzdelávacej organizácie. Tretia časť obsahuje otázky zamerané na zistenie miery vnímania potreby vzdelávania sa obyvateľov z hľadiska potenciálneho rozvoja regiónu a rastu životnej úrovne. Posledná časť dotazníka obsahuje otázky zamerané na definovanie hodnoty vzdelávania z hľadiska motivačných faktorov účasti na vzdelávaní. Dotazníky boli distribuované študentom prvého roka štúdia v dennej a externej forme v prvom a druhom stupni vysokoškolského štúdia na Fakulte manažmentu PU v Prešove, ako dominantnej vzdelávacej inštitúcie v regióne. Získaná databáza má celkovo 818 záznamov.

Z hľadiska dodržania predpokladov použitia log-lineárnej analýzy sme päťstupňovú škálu škálových otázok binarizovali do nových premenných, ich názvy sú o8II až o50II. Hodnota 1 je vymedzená pre zápornú, hodnota 2 pre kladnú odpoveď. Po frekvenčnej analýze týchto premenných sme vylúčili 22 z nich pre ich vysoko prevažujúce jednostranné odpovede. Pre potreby tohto článku vytrhneme z kontextu rozsiahleho výskumu iba niektoré súvislosti a ukážeme na nich použitie log-lineárnej analýzy.

**Príklad 1.** Uvažujme otázky:

**o28II:** Považujete ponuku služieb v oblasti vzdelávania v regióne za dostatočnú z hľadiska štruktúry a rozmanitosti ponuky študijných programov? 1 - nie, 2 - áno

**o32II:** Ak by ste sa mali na základe Vašich doterajších skúseností rozhodnúť, zvolili by ste si štúdium na tej istej vzdelávacej organizácii v tom istom študijnom programe z hľadiska jeho reálneho prínosu pre Váš osobnostný rozvoj a kvalifikačný rast v porovnaní s očakávaniami pri výbere vzdelávacej organizácie? 1 - nie, 2 - áno

**o2:** Váš vek: 1 - do 25 rokov, 2 - nad 25 rokov.

Pri analýze dvojíc týchto otázok (premenných) sme zistili významnú súvislosť odpovedí na otázku o32II s vekom, u ostatných dvojíc sa významná súvislosť nepreukázala. Pohľad do kontingenčných tabuliek každej z dvojíc premenných na rôznych úrovniach tretej (Tab. 7.) nás, aj z hľadiska dodržania predpokladov, motivoval analyzovať tieto otázky komplexnejšie, ako trojicu otázok log-lineárnou analýzou. Základnou motiváciou boli testy z rodiny chí-kvadrát. Analýza dvojíc premenných na rôznych úrovniach tretej zakaždým preukázala významnosť pri úrovni tretej premennej rovnej 2 a nevýznamnosť pri hodnote rovnej 1. Očakávali sme, že log-lineárna analýza vysvetlí túto situáciu.

Výsledok hierarchickej log-lineárnej analýzy (Tab 8.-9.) naše očakávania potvrdil. Najlepším modelom je tu model saturovaný, model s označením (o28II\*o32II\*vek). V modeli sú prítomné efekty prvého rádu, efekty druhého rádu aj efekt interakcie všetkých troch premenných. Žiaden úspornejší model nevysvetlí komplikovanú súvislosť vedúcu k rôznym významnostiam na rôznych úrovniach tretej premennej. Reziduá sú nulové, keďže model je saturovaný. Tabuľku rezíduí preto neuvádzame.

Tab. 7.

	vek	o32II	o28II		Tot.	o32II	vek	o28II		Tot.	o28II	vek	o32II		Tot.			
			1	2				1	2				1	2				
Count	1	1	11	29	40	1	1	11	29	40	1	1	11	83	94			
Exp. c.			9,4	30,6	40			8,9	31,1	40			9,4	84,6	94			
Count		2	83	279	362		2	1	13	14		2	1	25	26			
Exp. c.			84,6	277,4	362			3,1	10,9	14			2,6	23,4	26			
Count	2	Tot.	94	308	402	2	Tot.	12	42	54	2	Tot.	12	108	120			
Count		1	1	13	14		1	83	279	362		1	29	279	308			
Exp. c.			4,3	9,7	14			90,5	271,5	362			35,3	272,7	308			
Count		2	25	45	70		2	25	45	70		2	13	45	58			
Exp. c.			21,7	48,3	70			17,5	52,5	70			6,7	51,3	58			
Count		Tot.	26	58	84		Tot.	108	324	432		Tot.	42	324	366			
vek=1: $G^2=0,406$ ; $p=0,524$							o32II=1: $G^2=2,950$ ; $p=0,086$					o28II=1: $G^2=1,685$ ; $p=0,194$						
vek=2: $G^2=5,494$ ; $p=0,019$							o32II=2: $G^2=4,805$ ; $p=0,028$					o28II=2: $G^2=6,896$ ; $p=0,009$						

Tab 8. K-Way and Higher-Order Effects

K	df	Likelihood Ratio		Pearson		Number of Iterations
		Chi-Square	Sig.	Chi-Square	Sig.	
K-way and Higher Order Effects <sup>a</sup>	1	702,255	,000	970,988	,000	0
	2	10,842	,028	11,705	,020	2
	3	5,574	,018	4,846	,028	2
K-way Effects <sup>b</sup>	1	691,413	,000	959,283	,000	0
	2	5,268	,153	6,859	,077	0
	3	5,574	,018	4,846	,028	0

Tab 9. Step Summary

Step <sup>a</sup>	Effects		Chi-Square <sup>c</sup>	df	Sig.	Number of Iterations
0	Generating Class <sup>b</sup>	o2*o28II*o32II	,000	0	.	
	Deleted Effect	1 o2*o28II*o32II	5,574	1	,018	2
1	Generating Class <sup>b</sup>	o2*o28II*o32II	,000	0	.	

Zdroj: vlastné spracovanie v SPSS

Prepletené súvislosti troch premenných dokázané analýzou premietneme do výrazov, ktoré premenné zastupujú. U starších študentov je významná negatívna asociácia medzi spokojnosťou v ponuke služieb v oblasti vzdelávania v regióne a medzi preferenciou tej istej vzdelávacej inštitúcie, na ktorej študujú. Inými slovami, viacerí by si zvolili znova tú istú školu, lebo región im nič lepšie neponúka. U študentov, ktorí by si tú istú vzdelávaciu inštitúciu znova zvolili, je významná negatívna asociácia medzi vekom a spokojnosťou s ponukou služieb v oblasti vzdelávania, teda starší z nich sú menej spokojní s ponukou služieb regiónu v oblasti vzdelávania. U študentov spokojných s ponukou služieb v oblasti vzdelávania je významná negatívna asociácia medzi vekom a opätovným výberom tej istej vzdelávacej inštitúcie, teda menej starších študentov z nich by si znova zvolilo tú istú vzdelávaciu inštitúciu.

**Príklad 2.** Uvažujme názory:

**o42II:** Štúdium na vysokej škole v regióne je výhodou pri hľadaní si zamestnania v tomto regióne. 1 - nie, 2 – áno

**o45II:** Štúdium na vysokej škole prináša formu „sebarealizácie“ a je pre mňa dôležité z hľadiska môjho osobnostného rastu. 1 - nie, 2 – áno



**o49II:** Predpokladám, že po doštudovaní sa zamestnám v odbore môjho štúdia v našom regióne. 1 - nie, 2 – áno.

Skúmanie týchto troch vybraných premenných log-lineárnou analýzou (Tab. 10.-11.) vedie k optimálnemu výsledku (o45II\*o49II,o42II). V terminológii Cristante (2000) je to model štruktúry 2. V modeli sú prítomné efekty prvého rádu a jeden z efektov druhého rádu. Reziduálna analýza je veľmi priaznivá (Tab. 12) a aj 'Goodnes of fit test' ( $G^2=2,733$ ;  $p=0,435$ ) potvrdzuje vhodnosť optimálneho modelu aj vhodnosť použitia log-lineárnej analýzy z hľadiska dodržania predpokladov. Kontingenčné tabuľky (tu neuvedené, významnosť je len pre vzťah o45II-o49II, kde pre o42II=1 je  $G^2=230,469$ ;  $p=0,000$ ; pre o42II=2 je  $G^2=37,578$ ;  $p=0,000$ ) potvrdzujú výsledok získaný analýzou a sú užitočné pri interpretácii zistenej súvislosti.

Tab 10. K-Way and Higher-Order Effects

K	df	Likelihood Ratio		Pearson		Number of Iterations
		Chi-Square	Sig.	Chi-Square	Sig.	
K-way and Higher Order Effects <sup>a</sup>	1	772,136	,000	1018,992	,000	0
	2	269,052	,000	266,336	,000	2
	3	1,422	,233	,789	,374	6
K-way Effects <sup>b</sup>	1	503,083	,000	752,656	,000	0
	2	267,630	,000	265,547	,000	0
	3	1,422	,233	,789	,374	0

Tab 11. Step Summary

Step <sup>a</sup>	Effects		Chi-Square <sup>c</sup>	df	Sig.	Number of Iterations
0	Generating Class <sup>b</sup>	o42II*o45II*o49II	,000	0	.	
	Deleted Effect	1 o42II*o45II*o49II	1,422	1	,233	6
1	Generating Class <sup>b</sup>	o42II*o45II, o42II*o49II, o45II*o49II	1,422	1	,233	
	Deleted Effect	1 o42II*o45II	1,233	1	,267	2
		2 o42II*o49II	,384	1	,535	2
		3 o45II*o49II	266,625	1	,000	2
2	Generating Class <sup>b</sup>	o42II*o45II, o45II*o49II	1,806	2	,405	
	Deleted Effect	1 o42II*o45II	,927	1	,336	2
		2 o45II*o49II	266,319	1	,000	2
3	Generating Class <sup>b</sup>	o45II*o49II, o42II	2,733	3	,435	
	Deleted Effect	1 o45II*o49II	266,319	1	,000	2
		2 o42II	276,682	1	,000	2
4	Generating Class <sup>b</sup>	o45II*o49II, o42II	2,733	3	,435	

Tab. 12. Bunkové počtosti a reziduá

o42II	o45II	o49II	Observed		Expected		Residuals	Std. Residuals
			Count	%	Count	%		
1,00	1,00	1,00	89,000	18,7%	86,996	18,3%	2,004	,215
		2,00	7,000	1,5%	6,029	1,3%	,971	,395
	2,00	1,00	36,000	7,6%	37,038	7,8%	-1,038	-,171
		2,00	278,000	58,4%	279,937	58,8%	-1,937	-,116
2,00	1,00	1,00	12,000	2,5%	14,004	2,9%	-2,004	-,536
		2,00	,000	,0%	,971	,2%	-,971	-,985
	2,00	1,00	7,000	1,5%	5,962	1,3%	1,038	,425
		2,00	47,000	9,9%	45,063	9,5%	1,937	,289

Zdroj: vlastné spracovanie v SPSS

Výsledok premietnutý do výrazov, ktoré premenné zastupujú hovorí, že existuje silná významná pozitívna súvislosť medzi pocitom sebarealizácie a osobnostného rastu pri štúdiu na vysokej škole a predpokladom, že sa študent po ukončení školy zamestná vo svojom odbore aj vo svojom regióne. Ďalšou informáciou, vyplývajúcou z modelu ( $\phi_{45II} \cdot \phi_{49II}, \phi_{42II}$ ) je, že táto súvislosť je rovnaká nielen pre tých, ktorí svoj predpoklad opierajú aj o fakt, že miestom štúdia je ich región, ale aj pre tých, ktorí miesto štúdia z tohto hľadiska nepovažujú za dôležité. Poslednou informáciou je, že iná súvislosť uvažovaných premenných neexistuje.

Pre zaujímavosť sme ešte vybrali 4 už známe premenné, a našli sme model ( $\phi_{45II} \cdot \phi_{49II}, \phi_{45II} \cdot \phi_{28II}, \phi_{32II} \cdot \phi_{49II}, \phi_{32II} \cdot \phi_{28II}$ ). Podrobné výsledky ani interpretáciu neuvádzame. Grafickou interpretáciou je štvorec (viac Cristante, 2000).

## ZÁVER

Algoritmus hierarchickej log-lineárnej analýzy umožňuje medzi viac ako dvoma kvalitatívnymi premennými v jednom kroku nachádzať vzťahy, ktoré by sme pri použití jednoduchších metód zdĺhavo a komplikovane hľadali. Použitie metódy zároveň garantuje istotu, že nájdený vzťah je pre uvažované premenné najoptimálnejší.

## LITERATÚRA

1. AGRESTI, Alan, 2002. *Categorical data analysis. Second Edition*. University of Florida. Wiley Interscience, John Wiley & Sons, Inc., Hoboken, New Jersey. ISBN: 0-471-36093-7. Dostupné na internete: [http://books.google.com/books?id=hpEzw4T0sPUC&printsec=frontcover&dq=inauthor:Alan+Agresti&lr=&as\\_brr=0&hl=sk#v=onepage&q=&f=false](http://books.google.com/books?id=hpEzw4T0sPUC&printsec=frontcover&dq=inauthor:Alan+Agresti&lr=&as_brr=0&hl=sk#v=onepage&q=&f=false)
2. CRISTANTE, Francesca, 2000. *Log Linear analysis*. Open Distance Learning, Mathematical Psychology. Dostupné na internete: <http://www.uni-bonn.de/mathpsyc/doc/cristant/cristant.html>
3. GARSON, David, 2009. *Log-Linear, Logit, and Probit Models*. Dostupné na internete: <http://faculty.chass.ncsu.edu/garson/PA765/logit.htm>

## KONTAKT

**Mgr. Eva Litavcová**

**Katedra kvantitatívnych metód a manažérskej informatiky**

**Fakulta manažmentu, Prešovská univerzita**

**Konštantínova 16, Prešov 08001**

**[litavcov@unipo.sk](mailto:litavcov@unipo.sk)**

**Doc. Ing. Ivana Butoracová Šindleryová, PhD.**

**Katedra marketingu a medzinárodného obchodu**

**Fakulta manažmentu, Prešovská univerzita**

**Konštantínova 16, Prešov 08001**

**[ivkasindleryova@yahoo.com](mailto:ivkasindleryova@yahoo.com)**